Precoordination of diagnosis concepts for increased granularity in a clinical genomic ontology

1. INTRODUCTION

A. Clinical Genomics

Major healthcare advances have been made possible by the development of new pharmaceuticals, such as medicines for breast and prostate cancer, high blood pressure and gastric ulcers. The rate of progress is accelerating now as recent understandings in molecular biology are translated into novel approaches for the development of new therapies. Nevertheless, although the pharmaceutical and biotechnology industries invest over $45 billion each year in research and development [1, 2], the success rate in developing new therapies is still quite low. One important reason for this sub-optimal success rate has been the difficulty of validating, early enough in the drug discovery process, that both the biological target for a drug is relevant to the disease, and the drug itself is specific for the target.

In recent years, scientists have made significant progress in deciphering the underlying causes of disease at the molecular level. Through the Human Genome Project, researchers are now defining the molecular basis of disease processes in the human body at the level of the gene and biological pathway, discovering a large number of potential therapeutic targets [3, 4]. In addition, through advances in chemistry and other pharmaceutical discovery technologies, active molecules can be quickly discovered and pursued for potential drug development. The massive amount of information resulting from these efforts will eventually provide a vast compendium of

knowledge about disease. However, the near term result is that researchers are now awash in potential targets, most of them lacking sufficient information linking them to clinical relevance [5].

To fulfill the promise of these scientific advances, it is vital that researchers be able to select the most clinically important biological targets. To enable intelligent selection, potential targets must be annotated with information from multiple, essential sources. Specifically, scientists must efficiently link data from existing genomics databases with new molecular data generated from human tissue samples that are richly annotated with clinical information from the donor's medical profile.

"Clinical Genomics" is a new term for a multi-disciplinary science that is attempting a high-throughput correlation of molecular status, including differences in patterns of gene expression, with the characteristics of actual human disease, to accelerate the discovery and validation of targets for development of new diagnostics and therapeutics. A growing number of scientific studies are being published in which well-preserved and well-annotated human tissues have been prospectively collected and analyzed to understand the molecular basis of disease [6, 7]. To truly enable clinical genomics, researchers must have access to large numbers of high quality biological samples clinically annotated in a consistent, comparable and machine readable manner, which will result in a far more efficient and reliable ability to connect molecular and phenotypic information.

Access to case and sample level data from clinical care and surgical pathology is but the first step in a process of converting that information into a useful resource with the noted

characteristics of being structured, searchable and highly relevant to research. To facilitate this conversion, we have devised an ontology of clinical concepts specifically focused to support a community of genomic researchers working from a primarily molecular biology viewpoint. This ontology relies on a precoordinated controlled vocabulary to ensure the captured data are rendered useful to researchers working with large, electronic data sets.

B. Information needs of the clinical genomic research community

The process of defining a controlled vocabulary, in particular the task of assigning names to clinical concepts, must address the questions of the level of granularity to which the concepts are to be defined, and how that level is determined. Kim and Frosdick explore this question in the context of drug hierarchies [8], Hardiker and Kirby in nursing terminologies [9], and White, Kolar and Steindel in public health reporting [10]. Elkin, Brown and Chute discuss granularity in healthcare vocabularies in general, and summarize the imperative of understanding the purpose of a vocabulary well: "Any controlled vocabulary must have its purpose and scope clearly stated in operational terms so that its fitness for particular purposes can be assessed and evaluated." [11].

In the case of establishing biorepositories for genomic and proteomics research, these cornerstone questions are focused on the needs of users of the resulting sample and clinical data library: What level of concept granularity is needed by clinical genomic researchers? The answer is derived from two core information needs in the clinical genomic research community. First, there must be the ability to identify and distinguish clinical concepts at that level of granularity that makes them relevant to clinical medicine, molecular biology, genomic and biotechnology research and development. Second, there must be the ability to query and search on highly specific values for these concepts.

To meet these needs, we have undertaken the development of an ontology that is based on highly structured, fine-grained controlled vocabularies whose members are complete and unique in their representation of individual clinical concepts. The values of the controlled vocabularies are concepts which pathologists at donor institutions provide in pathology reports or which in-house pathologists provide during pathology verification processes, or which researchers specify when selecting samples from the biorepository. Table 1 lists twenty-eight diagnoses for Carcinoma of Lung that have been received or requested. This list demonstrates the level of concept granularity that must be distinguished for clinical genomic researchers. Note that several of these names are synonymous with more appropriate names of Adenocarcinoma of Lung with analogous morphologic subtyping. The "Carcinoma" names of the concepts are retained along with the preferred "Adenocarcinoma" names to support user preference.

However, there is no standard, broad, diagnosis-oriented ontology with this degree of morphology, histology and localization detail available within the healthcare information systems community. For example, SNOMED Clinical Terms (CT) [12] is the most comprehensive clinical terminology system available and represents an impressive expansion of clinical domain coverage from SNOMED Reference Terminology [13, 14]. SNOMED CT includes ten concepts for Carcinoma of Lung, six of which distinguished specific morphologic types: clear, giant, large, oat, small, and squamous cell carcinomas. Although a major step forward in concept detail, clinical genomic researchers need other and more complex carcinoma typing, such as non-small, neuroendocrine, and large and small cell carcinomas.

Building the bridge between clinical medicine and molecular biology requires a language that 'works' for both the healthcare practitioner and the genomic researcher. The urgency of this need, and the lack of suitable standard vocabularies to meet this need, have lead us to develop the ARTS™ ontology.

2. MEETING THE INFORMATION NEEDS OF GENOMIC RESEARCHERS

A. Clinical knowledge representation for genomic research

ARTS is an ontology of clinical concept controlled vocabularies, codes, and relationships that provides a unified and standardized language with which to represent and communicate clinical information within a biorepository facility, and between the biorepository, its sourcing medical centers partners, and the researcher. The content domains addressed by ARTS include diagnoses, tissue/body systems, surgical procedures, donor demographics, attributes of neoplastic and non-neoplastic diseases, attributes of donor treatment, and attributes of donor past medical history.

Clinical knowledge representation is based on the approach that clinical concepts are units of thought that can be labeled with a name and a code. A concept name is a symbol used to represent the concept that looks like one or more words. A concept code is a symbol used to represent the concept that does not look like words. There is a one-to-one relationship between concepts and codes: every concept is represented by one code, and that code represents only one concept. Codes do not change and are not recycled when operations such as Split and Merge are performed on concepts to reflect changing understanding of user requirements and genomic science. In contrast, concept names may be changed and there can be multiple names for a

single concept. As in many biomedical taxonomies, one of the concept names is designated as the preferred, and the other names are synonymized to it (be they called 'synonyms'. 'strings', 'terms', 'descriptors', or 'descriptions').

B. Precoordination of diagnosis concepts

The data element of greatest importance in clinical genomics, and the knowledge domain least satisfactorily provided for by the healthcare information systems industry, is that of the clinical diagnosis. Representation of clinical diagnoses is a prime focus of our applied informatics research and development effort, and is the subject of this report.

As mentioned, the requirements for the ARTS ontology are driven by its users – the content and level of specificity defined by the suppliers and consumers of clinical data. When there is discordance between these two interests, the language of research clients takes precedence over that of clinical care delivery as abstracted from pathology reports or other medical record documents. Our approach to meeting these requirements in the domain of diagnoses is to define, de novo, diagnosis concepts at the desired granularity, and to construct names and codes for these concepts from simpler clinical concepts. The new diagnosis concept definitions are precoordinated. That is, they are created prior to their availability to users. The constituent concepts are treated as building blocks in the precoordinated construction of the more complex concepts. The building block concepts are analogous to atoms that are bonded together into stable molecules prior to their interactions in the real world.

Knowledge representation methodologies have been widely discussed in the medical informatics literature, including comparison of precoordination and postcoordination of simpler concepts

into more complex. In these discussions, precoordination has also been labeled 'enumerative' concept definition. 'Compositional' concepts have been defined to mean that concepts have been constructed from other concepts, whether 'compositional' is restricted to only postcoordinated concept definition [15, 16], or it is applied to both precoordinated and postcoordinated [11, 17, 18]. The constituent concepts may themselves be simple (or 'atomic', 'elementary', or 'primitive') or previously precoordinated.

These comparisons have sometimes presented precoordination and postcoordination as opposing methodologies. Precoordination has been portrayed as an inferior approach because of anticipation of an uncontrolled and unmanageable proliferation of concepts [10, 18, 19, 20]. The "SNOMED Record Abstracting Guidelines" from the Virginia-Maryland Regional College of Veterinary Medicine states this view succinctly: "If nomenclature is to pre-coordinate all potential medical circumstances in such elaborate detail, the number of concepts will burgeon, kudzu-like, into an impenetrable mass" [21]. It should be noted that postcoordination is subject to an analogous weakness, multiplicity of combinations of independent concepts being used to represent more complex concepts, as well as nonsensical association of concepts [10, 11, 17, 20, 22 - 24].

Spackman and Campbell delineate four approaches to knowledge representation using compositional conceptual models (CCM) [20]. Three of the approaches, which are exemplified by the evolution of SNOMED, have weaknesses that are discussed. The ultimate approach, they conclude, is implementation of a description logic language (also called 'frame-based representation' [25]) which has since been incorporated into SNOMED CT.

The approach of precoordination of diagnosis concepts might be categorized by Spackman and Campbell as Multi-axis Composition, CCM-2. They criticize this approach for leaving implicit or undefined the specific attribute for which the desired compositional component is the value to be used, and for limiting the expressiveness of the representation because the range of attributes of interest exceeds the number of orthogonal axes by which concepts are grouped. It can be inferred that the authors are analyzing postcoordination in this critique: they would probably add, if precoordination was the concept definition process being reviewed, the failing of proliferation of concepts to support a broad clinical domain.

Tension between precoordination and postcoordination methodologies can be greatly reduced by bringing an explicit statement of the purpose of the vocabulary into the discussion. Precoordination becomes unmanageable when the content domain is broad, such as support of an electronic medical record. For such purposes, postcoordination of vocabulary is a necessary complement. The "SNOMED Record Abstracting Guidelines" is again concise and precise in making this point. Discussing the role of precoordination in ICD-9, CPT, and early versions of SNOMED, Wilcke et al. write: "In each of these cases, the nomenclature's intended use both limited the content required and implied a semantic context that added meaning to the vocabulary. … But the wider applications – particularly comprehensive medical records – will ultimately require more than the endless addition of new precoordinated concepts." [21]

Indeed, there are in the medical informatics literature analyses that go beyond encouraging postcoordination as a 'better' methodology than precoordination, to an appreciation that both approaches to complex knowledge representation are widely used and need to be accommodated

in information systems development [17, 26, 27], or that the two approaches may even be synergistic [16], or that precoordination may even control terminology proliferation better than postcoordination [9, 22].

Our context is the need for vocabulary for a set of relatively focused content domains – users require specific clinical concepts for particular documentation, communication and retrieval tasks. In each domain, the concept attributes of interest are explicitly stated and the range of attributes conforms to the user requirements. Excessive and irrelevant vocabulary proliferation is obviated by the approach of adding complex concepts as they are needed, and there is no attempt to comprehensively preload an encyclopedic ontology. Precoordination allows definition at the level of granularity needed to distinguish and retrieve concepts of interest to researchers.

C. Implementation of precoordinated diagnosis concepts

For diagnosis concept names, syntax rules have been created based on SNOMED concept names and the cancer synoptic reports developed by the College of American Pathologists [28]. The language is familiar to and easily understood by users of our information systems, where it replaces idiosyncratic, parochial, and/or telegraphic diagnosis language often found in clinical documents. This normalized diagnosis controlled vocabulary provides explicit, concise, complete and predictable names across donor institutions and clients, for both data entry and query parameter setting. Figure 1 is a screen shot of the data entry screen used during pathology report abstraction (the recording of relevant pathology findings into structured data fields), showing a portion of the diagnosis controlled vocabulary pick list.

The general syntax for diagnosis concept names is a comma delimited string comprising Disease or Morphology, Site, Morphology Subtype, other Modifiers. Examples of three diagnosis name syntax rules are:

- identification of tumor specimens as metastases or recurrent tumors is very important, and that status is explicitly included in the diagnosis concept name
- identification of tumor specimens as invasive or infiltrative is also important, however this status is not distinguished from the primary tumor in the diagnosis name, but is documented through the difference in the values of data elements for the tissue of origin of the disease and the site of finding of the specimen
- identification of specimens as residual tumor is less important and is not recorded as different from the primary lesion

Codes for the new diagnosis concepts are created by concatenation of up to three building block concept codes, separated by a "^" delimiter, and for which normalization rules have also been created. A SNOMED Disease or Finding axis concept identifier (or a local extension identifier) is required as the first component of the diagnosis concept code, followed by an optional Morphology concept identifier, which is then followed by option General or Topographic concept identifiers. Table 2 presents the concept names and codes for six morphologic types of the Carcinoma of Lung supported by ARTS.

Non-neoplastic diagnoses are also represented with precoordinated concepts names and codes, as presented in Table 3. In these diseases, morphologic and other typing is often included in the

concept code to facilitate automated query across diagnosis groups for attributes such as chronicity, presence of inflammation or infection, and atypia.

There are gaps in SNOMED-CT for the concept specificity and granularity required for clinical genomic research. As in many vocabulary development efforts [14], these gaps by creating local concept extensions. These are used as building blocks, identically to those of SNOMED, in precoordinated concept construction. Table 4 presents examples of concept extensions that supplement concepts used from SNOMED-CT. The local extensions are expected to be replaced as these concepts become available in SNOMED.

D. Detecting and remedying vocabulary gaps and redundancies

One of the metrics of good controlled vocabulary practice is the degree to which the vocabulary covers the concepts in its domain or scope [22 - 24]. Because the creation of diagnosis concepts is triggered by data collection or user query, the discovery by users of gaps in diagnosis domain coverage provides an on-going pipeline of candidates for new precoordinated concepts. The detection and remedying of these gaps are managed through a process called Other Code Edits (OCE), through which users report their discovery of gaps during data entry.

The foundation of OCE is the inclusion of the concept "Other" in each controlled vocabulary domain. Users are trained to select "Other" when the domain value they desire is not evident in the controlled vocabulary pick list for the data field of interest. When "Other" is selected in any clinical data entry application, a text entry window is activated into which the user is required to type something, hopefully the name of the concept they would have liked to pick from the

controlled vocabulary. The code for "Other" and the desired text is posted to the donor data table appropriate for the domain.

The code for "Other", the desired text, and data element and user identifiers are also posted to a special "Others" table. This table is accessed by an OCE editing application in which the desired text is displayed along with source documents (such as the ASCII pathology report for "Others" entered during pathology report abstraction), and the controlled vocabulary list appropriate for the data element. The editor reviews the desired text and if it is found that the correct concept is already available, the OCE application updates the "Other" code in the donor data table with the code for the concept.

If the desired concept is not already available, the editor sets a status flag for the "Other" text to denote that a new concept definition is required, that the "Other" needs off-line review, or that the text entered is inappropriate. New concepts are queued by domain for inclusion in periodic controlled vocabulary and code updates. "Others" needing review are printed with their source documents for regular multidisciplinary terminology reviews. Inappropriate text is referred to the product manager of the application through which the "Other" text was entered. On a monthly basis, product managers are also given a list of "Other" usage in their applications along with feedback to be discussed with their users concerning how to find correct concept names and/or appreciation for vocabulary gaps discovered through the OCE process.

3. DISCUSSION

The use of precoordinated diagnosis concepts described in this report supports a vibrant communication between the biorepository, medical center partners who provide tissue and data, and academic and commercial research partners who use tissue and data. Potential drawbacks in the concept precoordination approach have not been an obstacle. In our opinion, post-coordination of concepts to document complex clinical ideas such as diagnoses offers much greater opportunity for redundant, user-crafted representations and irrelevant concept combinations than does user-directed precoordination of controlled vocabulary.

Is there a role for post-coordination in a clinical genomics data collection system? Absolutely - particularly in domains where irregular permutations of multiple modifiers to clinical concepts are required. For example, an Anatomical Compass is being developed to improve tissue and organ data collection. The Compass does not address tissue granularity, such as enumeration of specific long bones or major peripheral arteries, which is supported in the current Tissue ontology. Instead, the Anatomical Compass allows specification of aspects of Tissue concepts, such as left – right, lateral – medial, anterior – posterior, distal – proximal, etc. Any number of these independent attributes may be toggled and associated with a selected Tissue during data entry.

Plans are also being made to migrate ARTS to a description logic methodology to support future growth. Description logic is a powerful and flexible approach for the creation of complex knowledge representation systems, and has been adopted by several healthcare information systems and vocabulary developers [9, 25, 29].

Description logic is entirely compatible with the approach of precoordinating complex concept definition. The level of specificity of clinical concepts will continue to be defined by the needs of the clinical genomic research community and diagnosis concepts of appropriately granular morphology detail will be distinguished. Description logic will allow a more robust ontology by facilitating complex concept creation and by expanding the quality and quantity of relationships between concepts.

SNOMED provides excellent building blocks with which to name and code the specific concepts important to our business. Other standard ontologies and taxonomies are being reviewed, and portions of them may be incorporated into ARTS. However, neither SNOMED nor other current knowledge representation systems necessarily need to provide individual concepts at the level of specificity we require. They should continue reviewing feedback they receive and add concepts that support the information needs of more general users.

4. CONCLUSIONS

Evaluating the efficacy of the controlled vocabulary, coding systems and relationship definition that constitute a specific ontology requires explicit specification of the purpose of the ontology, to its users and to the enterprise or business adopting it. Elkin et al., in discussing medical record problem lists, ask and then answer the key question in ontology requirements specification: "If composition causes such angst, why do it? Why not ignore this functionality? … Users demand the ability to form problem statements that represent the concepts of their practice." [18]

The new field of clinical genomic research requires structured clinical data collection and manipulation predicated on the identification and electronic representation of precise, fine-grained and distinct concepts. Precoordination of complex diagnosis concepts from simpler 'building block' concepts meets this need.

Acknowledgements

The authors are grateful to Alan J. Buckler, PhD for his thoughtful review of this manuscript.

David B. Aronow and Martin L. Ferguson are both employees and shareholders of Ardais Corporation.

References

1. Biotechnology Industry Organization, www.bio.org; Biotechnology Industry Statistics. http://www.bio.org/er/statistics.asp.
2. Pharmaceutical Research and Manufacturers of America, www.PhRMA.org; Research and Development Overview. http://www.phrma.org/issues/researchdev/.
3. The Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409: 860 – 921.
4. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. Science 2001; 291: 1304-1351.
5. Lehman Brothers and McKinsey & Co. "The Fruits of Genomics: Drug Pipelines Face Indigestion Until The New Biology Ripens". January 2001.
6. Alizadeh AA, Elsen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000; 403: 503 – 511.
7. Dhanasekaran SM, Barrette TR, Ghosh D, et al. Delineation of prognostic biomarkers in prostate cancer. Nature 2001: 412: 822-826.
8. Kim JM, Frosdick P. Description of a drug hierarchy in a concept-based reference terminology. Proc AMIA Symp 2001: 314-318.
9. Hardiker N, Kirby J. A compositional approach to nursing terminology. Stud Health Technol Inform 1997; 46: 3-7.
10. White MD, Kolar LM, Steindel SJ. Evaluation of vocabularies for electronic laboratory reporting to public health agencies. J Am Med Inform Assoc 1999; 6(3): 185-194.

11. Elkin PL, Brown SH, Chute CG. Guidelines for health informatics: controlled health vocabularies – vocabulary structure and high-level-indicators. MedInfo 2001; 10(Pt 1) :191-195.

12. Spackman KA, et al, eds. SNOMED Clinical Terms. Northfield IL: College of American Pathologists, 2002.

13. Spackman KA, et al, eds. SNOMED RT: Systematized Nomenclature of Medicine, Reference Terminology. Northfield IL: College of American Pathologists, 2000.

14. Wang AY, Sable JH, Spackman KA. The SNOMED Clinical Terms development process: refinement and analysis of content. Proc AMIA Symp 2002: 845-849.

15. NHS Information Authority Clinical Terminology Service. "Principles of Terminology". http://www.nhsia.nhs.uk/terms/seminar/03Principlesv2.0.ppt. 2002.

16. Rassinoux A-M, Miller RA, Baud RH, Scherrer J-R. Compositional and enumerative designs for medical language representation. Proc AMIA Symp 1997: 620-624.

17. Dolin RH, Spackman KA, Markwell D. Selective retrieval of pre- and post-coordinated SNOMED concepts. Proc AMIA Symp 2002: 210-214.

18. Elkins PL, Tuttle M, Keck K, Campbell K, Atkins G, Chute CG. The role of compositionality in standardized problem list generation. MedInfo 1998; 9 Pt 1: 660-664.

19. Brown PJB. Principles of post-coordination concept construction and retrieval. Proc AMIA Symp 2001: 869.

20. Spackman KA, Campbell KE. Compositional concept representation using SNOMED: towards further convergence of clinical terminologies. Proc AMIA Symp 1998: 740-744.

21. Wilcke JR, Livesay P, Zimmerman K. SNOMED Record Abstracting Guidelines. Virginia-Maryland Regional College of Veterinary Medicine. 2002. http://snomed.vetmed.vt.edu/guidelines/main.htm

22. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998; 37: 394-403.

23. De Keizer NF, Abu-Hanna A. Understanding terminological systems II: experience with conceptual and formal representation of structure. Methods Inf Med 2000; 39: 22-29.

24. Ingenerf J, Giere W. Concept oriented standardization and statistics oriented classification: continuing the classification versus nomenclature controversy. Methods Inf Med 1998; 37: 527-539.

25. Oliver DE, Shahar Y, Shortliffe EH, Musen MA. Representation of change in controlled medical vocabularies. Artif Intell Med 1999 Jan;15(1):. 53-76.

26. HL7 Vocabulary Technical Committee. Background. http://www.hl7.org/Special/committees/Vocab/vocab.htm.

27. Rector AL. Clinical terminology: why is it so hard? Methods Inf Med 1999; 38(4-5): 239-52.

28. College of American Pathologists. Cancer Protocols. http://www.cap.org/cancerprotocols/protocols_index.html.

29. Spackman KA, Dionne R, Mays E, Weis J. Role grouping as an extension to the description logic of Ontylog motivated by concept modeling in SNOMED. Proc AMIA Symp 2002: 712-716.

Figure 1. Data entry screen for pathology report abstraction

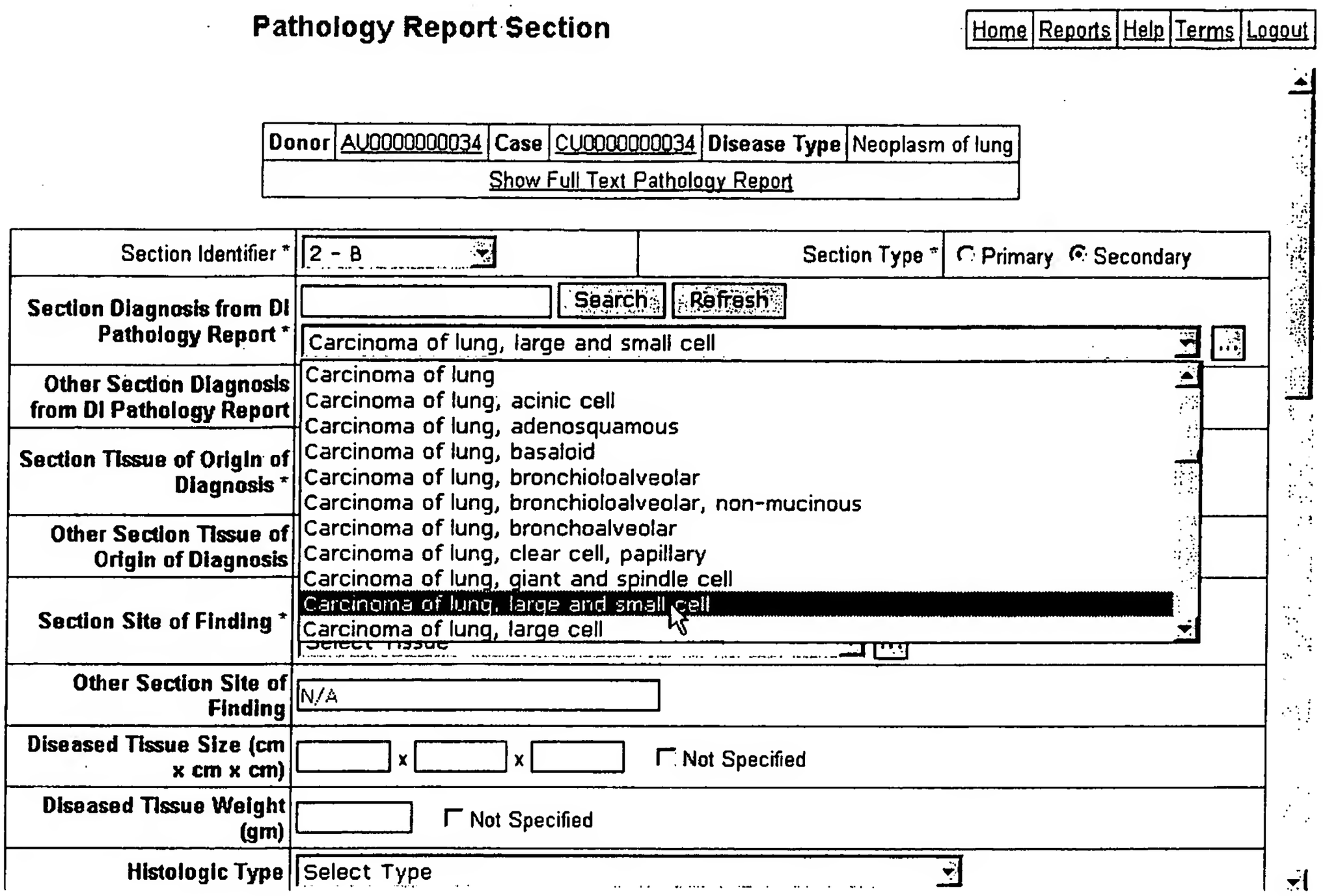

Table 1. Carcinoma of Lung diagnoses

Carcinoma of lung

Carcinoma of lung, acinic cell

Carcinoma of lung, adenosquamous

Carcinoma of lung, basaloid

Carcinoma of lung, bronchioloalveolar

Carcinoma of lung, bronchioloalveolar, non-mucinous

Carcinoma of lung, bronchoalveolar

Carcinoma of lung, clear cell, papillary

Carcinoma of lung, giant and spindle cell

Carcinoma of lung, large and small cell

Carcinoma of lung, large cell

Carcinoma of lung, large cell, metastatic

Carcinoma of lung, large cell, neuroendocrine

Carcinoma of lung, large cell, neuroendocrine, metastatic

Carcinoma of lung, mucinous

Carcinoma of lung, neuroendocrine

Carcinoma of lung, non-small cell

Carcinoma of lung, non-small cell, metastatic

Carcinoma of lung, oat cell

Carcinoma of lung, papillary

Carcinoma of lung, pleomorphic

Carcinoma of lung, sarcomatoid

Carcinoma of lung, small cell

Carcinoma of lung, small cell, metastatic

Carcinoma of lung, spindle and squamous cell

Carcinoma of lung, spindle and squamous cell, metastatic

Carcinoma of lung, squamous cell

Carcinoma of lung, squamous cell, metastatic

Table 2. Examples of diagnosis concept names and codes for morphologic types of Carcinoma of Lung

Carcinoma of lung, large cell

is coded 126713003^22687000^

= Neoplasm of lung, Large cell carcinoma

Carcinoma of lung, large cell, metastatic

is coded 126713003^22687000^8707003

= Neoplasm of lung, Large cell carcinoma, Metastatic

Carcinoma of lung, large and small cell

is coded 126713003^21326004^

= Neoplasm of lung, Combined small cell carcinoma

Carcinoma of lung, large cell, neuroendocrine

is coded 126713003^128628002^

= Neoplasm of lung, Large cell neuroendocrine carcinoma

Carcinoma of lung, large cell, neuroendocrine, metastatic

is coded 126713003^128628002^8707003

= Neoplasm of lung, Large cell neuroendocrine carcinoma, Metastatic

Carcinoma of lung, non-small cell

is coded 126713003^128632008^

= Neoplasm of lung, Non-small cell carcinoma

Table 3. Examples of diagnosis concept names, codes and code descriptors for Non-Neoplastic Diseases

Pyelonephritis, acute

is coded 36689008^4532008^53737009

= Acute pyelonephritis, Acute inflammation, Acute

Pyelonephritis, acute and chronic

is coded 45816000^75889009^CA00223G

= Pyelonephritis, Acute and chronic inflammation, Acute and chronic

Pyelonephritis, chronic

is coded 63302006^84499006^90734009

= Chronic pyelonephritis, Chronic inflammation, Chronic

Pyelonephritis, granulomatous

is coded 63302006^6266001^

= Chronic pyelonephritis, Granulomatous inflammation

Table 4. Examples of concept extensions to SNOMED-CT to provide missing building-block concepts

CA00010G	CAP College of American Pathologists
CA00013T	Peripheral artery
CA00015D	Neoplasm of cardiovascular system
CA00227C	Moderately to well differentiated
CA00374P	Gastric mass excision
CA00422C	AJCC Sixth Edition
CA00452D	Small airway disease
CA00501C	Medications on Admission
CA00596C	Lymphoid infiltrate
CA079954	brca1
CA079953	brca2
CA079915	S-phase